Reference: *Biol. Bull.* **183**: 123–137. (August, 1992)

Characterization of a Cystine-Rich Polyphenolic Protein Family from the Blue Mussel *Mytilus edulis* L.[1]

LESZEK M. RZEPECKI[1], KAROLYN M. HANSEN, AND J. HERBERT WAITE

College of Marine Studies, University of Delaware, 700 Pilottown Road, Lewes, Delaware 19958

Abstract. Marine bivalve mollusks synthesize, in the phenol and accessory glands of the foot, proteins that integrate the post-translationally hydroxylated amino acid 3,4-dihydroxyphenylalanine (DOPA) into their primary sequence. These polyphenolic proteins serve as structural and adhesive components of the byssal threads which form the extraorganismic holdfast. One family of byssal precursors, previously characterized in a number of mytiloid species, consists of proteins between 70–130 kDa containing 8–18 mol % DOPA. The high molecular weight precursor isolated from the foot of the blue mussel (*Mytilus edulis* Linnaeus, 1758) is here designated as Mefp-1. We now present evidence for the occurrence, in *M. edulis*, of a second, structurally unrelated, family of DOPA proteins (Mefp-2) of about 42–47 kDa. These novel proteins contain 2–3 mol % DOPA and, in startling contrast to Mefp-1, are also enriched in the disulphide-containing amino acid cystine (6–7 mol %). Consideration of the amino acid compositions of Mefp-1 and 2 and of the terminal adhesive plaques of byssal threads suggests that Mefp-2 makes up about 25% of plaque protein, whereas Mefp-1 content is about 5%. The Mefp-2 family exhibits electrophoretic microheterogeneity, but members share similar N- and C-terminal amino acid sequences. Analysis of peptides isolated after tryptic hydrolysis suggests that the primary sequence of Mefp-2 is tandemly repetitive, with at least three types of motif. The sequence degeneracy of the motifs is greater than in Mefp-1. Mefp-2 has minimal sequence homology with known structural proteins and may be a structural element of the plaque matrix.

Received 27 March 1992; accepted 20 May 1992.

Abbreviations: DOPA, 3,4-dihydroxyphenylalanine; Mefp-1 & -2, *Mytilus edulis* foot protein 1 and 2; CAMC; S-carbamidomethylcysteine; COMC, S-carboxymethylcysteine; NBA, Na/Borate/Ascorbate buffer; DTT, dithiothreitol; HPLC, high pressure liquid chromatography; EDTA, ethylenediaminetetraacetic acid; SDS- or AU-PAGE, sodium dodecylsulphate or acetic acid-urea polyacrylamide gel electrophoresis; IEF, isoelectric focussing; PAS, periodic acid-Schiff stain; NBT, nitroblue tetrazolium.

[1] Corresponding author.

Introduction

Mussels make a secure adhesive holdfast (byssus) that bonds tightly to wet and irregular surfaces without the extraordinary preparative treatments required when humans try their hand at similar tasks (Waite, 1987). The byssus, a collagenous extraorganismic tendon, is a bundle of threads that terminate in an adhesive plaque. The byssus is synthesized by a muscular organ, the mussel foot (see schematic diagram in Fig. 1). The thread coalesces from the secretions of various glands lining the ventral groove of the foot, while the plaque forms at the distal depression. The collagenous core of the thread is secreted by the collagen gland (Vitellaro-Zuccarello, 1980), and a protective varnish, made of an *o*-diphenolic resin coupled with a curing enzyme, is secreted by the accessory gland and applied to the thread cortex (Brown, 1952; Pujol, 1967; Vitellaro-Zuccarello, 1981). The resin is known to be a 130 kDa, tandemly repetitive protein (here called Mefp-1, *Mytilus edulis* foot protein), which incorporates into its primary structure a high proportion of 3,4-dihydroxyphenylalanine (DOPA), synthesized post- or co-translationally through the action of a tyrosyl hydroxylase (Waite, 1983; Benedict and Waite, 1986a). The curing enzyme is a catecholoxidase, activated after secretion of the resin, which converts the peptidyl DOPA residues into peptidyl DOPA-quinone, a probable cross-linking agent during sclerotization (Waite, 1990; Rzepecki and Waite, 1991, for reviews).

The plaque is formed from glandular secretion at the distal depression, and the main gland involved appears to be the phenol gland, so called because of its intense affinity for phenol and histological reagents (Tamarin *et al.*, 1976). Antibodies to Mefp-1 bind strongly to phenol and accessory glands (Benedict and Waite, 1986a), and Mefp-1 has been isolated from dissected phenol glands (Waite, 1983). Mefp-1 has also been identified as the adhesive at the interface between plaque and substratum (Benedict and Waite, 1986a), and its physicochemical properties are consistent with such a role (Filpula *et al.*, 1990).

Little is known, however, about the components of the plaque matrix. We have now characterized another potential ingredient of phenol gland secretion, Mefp-2, which may be a structural element of plaque matrix. Mefp-2 incorporates some DOPA, but is distinct from Mefp-1 in composition, sequence, and physicochemical properties, and hence probably serves a different role in byssal structure.

Materials and Methods

Purification of DOPA proteins

Cultured blue mussels (*Mytilus edulis* L.) were obtained via Rehoboth Seafood Market (Rehoboth Beach, Delaware) after being shipped live from Maine on ice. Feet were excised within two days of original shipment and stored in 30-g lots at −70°C.

The DOPA proteins were extracted and purified, as described by Pardo *et al.* (1990) and modified by Rzepecki *et al.* (1991). Briefly, 30 g of mussel feet were homogenized in a Waring blendor in 300 ml 0.7% perchloric acid, and the homogenate was sedimented at 31,000 × *g* to remove particulate matter. The DOPA proteins precipitated upon the addition of acetone to 66% final volume. The precipitate was resolubilized in 5–7 ml 5% acetic acid and then fractionated, in the same buffer, on a 2.5 × 90 cm column of Sephadex G-200 or G-150 (Pharmacia LKB Biotechnology Inc., Piscataway, New Jersey). The eluate was monitored at 280 nm. The final purification of Mefp-2 was achieved as follows. The lyophilized Sephadex fractions were chromatographed in 5% acetic acid on a 2.5 × 90 cm column of Sephacryl S-300 or S-400 (Pharmacia) monitored at 280 nm. This step was followed by high pressure liquid chromatography (HPLC) on a 0.7 × 25 cm semi-preparative Brownlee Aquapore RP-300 (C_8) reverse phase column (Rainin Instrument Co., Woburn, Massachusetts) with a Rabbit HPX solvent delivery system (Rainin); monitoring was at 280 nm with a Model 116 dual wavelength detector (Gilson Medical Electronics Inc., Middleton, Wisconsin). The acetonitrile/water solvent gradients contained 0.1% trifluoracetic acid.

In all of the chromatography procedures, the separation of Mefp-1 and Mefp-2 was assessed by gel electrophoresis (see below). Proteins were lyophilized after HPLC purification and stored dry at −20°C. DOPA concentrations were determined by the Arnow (1937) method (Waite and Tanzer, 1981), and the protein concentrations were determined according to Bradford (1976), with a Bio-Rad (Richmond, California) reagent kit.

The localization of Mefp-2 was studied in specimens of *M. edulis* that were collected locally (Roosevelt Inlet, Lewes, Delaware), kept in flowing seawater at 12–15°C, and allowed to deposit byssal threads on to plexiglass plates. In one set of experiments, some mussel feet were excised and serially sectioned, as illustrated in Figure 1a. The acid-urea soluble protein was extracted from each section by homogenization in a small glass homogenizer containing 200–300 μl of 5% acetic acid, 8 *M* urea. The particulate matter was sedimented at 13,000 × *g* for 5 min and was then analyzed for acid-soluble proteins by electrophoresis. In a second set of experiments, byssi were collected 24 h after their synthesis and rinsed in water; the plaques and threads were then separated with a scalpel. About 110–140 mg and 150–175 mg wet weight of plaques and threads, respectively, were homogenized, with a little silica powder (Silica Gel G, 250 μm Analtech, Newark, Delaware) to provide abrasion, in either 400 μl 5% acetic acid with 8 *M* urea, or in 200 μl of the sample buffer used for SDS-PAGE (Laemmli, 1970); the particulate matter was removed as above. Polyclonal anti-Mefp-2 rabbit antibodies were contractually prepared by Cambridge Research Biochemicals (Cambridge, Great Britain), and were used to determine the immunoreactivity of thread and plaque extracts and control Mefp-2 (Blake *et al.*, 1984). Dot blots on nitrocellulose were exposed for 1 h to a 10^{-5} dilution of serum, or a 10^{-4} dilution of pre-immune serum. Binding was detected with goat anti-rabbit IgG-coupled alkaline phosphatase (Boehringer-Mannheim, Indianapolis, Indiana).

S-Alkylation of Mefp-2

S-Alkylation of Mefp-2 to *S*-carbamidomethyl-Mefp-2 (CAMC-Mefp-2) or *S*-carboxymethyl-Mefp-2 (COMC-Mefp-2) was effected after the cystine residues had been reduced by a modification of a procedure of Hollecker's (1990). Stock solutions of Mefp-2 (10 mg/ml) were freshly prepared in H_2O. To 80 μl Mefp-2 was added 20 μl of 0.1 *M* EDTA, pH 7.0, which caused precipitation. After the addition of 400 μl of NBA buffer (0.5 *M* boric acid, 0.5 *M* ascorbic acid, brought to pH 8.0 with 6 *N* NaOH), the precipitate redissolved: 20 μl of fresh 1 *M* dithiothreitol (DTT) together with 1.15 ml of 8 *M* urea were then added, and the reaction mixture was incubated for 40 min at room temperature. After the reduction, the cyteines were

S-alkylated by the addition of fresh iodoacetamide or iodoacetate (300 μl of a 0.25 *M* stock in a 1:4 dilution of the NBA buffer for iodoacetamide, or, for iodoacetate, undiluted NBA buffer containing 0.6 μl of 6 *N* NaOH per mg iodoacetate), and the reaction mixture was incubated for a further 30 min. The reaction was terminated by the addition of 20 μl glacial acetic acid, and the proteins were immediately separated from the low molecular weight solutes by HPLC, as described. Control alkylations of Mefp-2 in buffers with urea, but no DTT, were also performed. Protein fractions were lyophilized, redissolved in H_2O, and stored at 4°C for up to 1 week before use.

Peptide preparation and purification

Native Mefp-2 or its alkylated derivatives were incubated (in 100 m*M* Tris, 100 m*M* ascorbate, pH 7.5, for 7 h at room temperature) with trypsin (Boehringer Mannheim, Indiana), at a trypsin to protein weight ratio of 1:25 or 1:50 for native and alkylated Mefp-2, respectively. Mefp-2 concentrations were about 1 mg/ml. The incubation was terminated by the addition of glacial acetic acid to 5% final volume, and the peptides were resolved directly by HPLC on a 0.46 × 25 cm Microsorb C_{18} reverse phase column (Rainin), as described above. In an effort to recover the DOPA peptides specifically, Mefp-2 (0.5 mg/ml) was treated with trypsin in 400 μl of a 1:2 dilution of the above buffer for 6 h, then loaded on a small column containing 0.4 ml of phenyl boronate agarose (PBA-30, Lot # JM-2134D with 37.8 μMoles boron/ml; Amicon Corp., Danvers, Massachusetts) which had been equilibrated in a 1:5 dilution of the above buffer. Bound peptides were washed twice with 0.4 ml of the diluted buffer, twice more with 0.4 ml of deionized water, and eluted with three washes of 5% acetic acid. Eluted peptides were separated by HPLC.

Amino acid analysis and peptide sequencing

The amino acid compositions of the hydrolyzed proteins and peptides were determined on a Beckman System 6300 Autoanalyzer (Beckman Instruments, Palo Alto, California) using the ninhydrin reaction for detection (Waite, 1991). Polypeptides were usually hydrolyzed by the rapid method of Tsugita *et al.*, in 5 *N* HCl with 8% trifluoracetic acid and 8% phenol *in vacuo* at 158°C for 22 min, except that hydrolysis was effected in the bulk phase, rather than the vapor phase as in the original method. Unfortunately, the high concentrations of phenol required for quantitative yields of DOPA caused the reaction of phenol and cystine to give an unidentified product which co-migrated with histidine on the amino acid analyzer. Consequently, the amino acid compositions of native (non-alkylated) Mefp-2 were determined after hydrolysis in 6 *N* HCl, for 24 h at 105°C, in the presence or absence of 10% phenol. DOPA, tyrosine, serine, and threonine values were recovered from phenolic hydrolysates, while cystine and histidine values were obtained from aphenolic hydrolysates; we deduced a corrected composition from both data sets using the lysine value as a common factor. In the case of reduced and alkylated Mefp-2, alkylated cysteine was recovered after hydrolysis as *S*-carboxymethylcysteine (*S*-COMC) which co-migrated on the amino acid analyzer with 3-hydroxyproline. Because native Mefp-2 contained no 3-hydroxyproline, the content of *S*-COMC was simply determined by applying a correction factor of 2.23 to the reported 3-hydroxyproline values. This correction factor was empirically determined by running a concentration series of *S*-COMC standards (Sigma) on the analyzer. Peptides and proteins were sequenced on a microsequenator (Porton Instruments, Tarzana, California) using automated Edman degradation. PTH-amino acid derivatives were resolved and quantitated by HPLC as previously described (Waite, 1991). Additional N-terminal analysis was performed on 4.6 mg/ml Mefp-2 in a buffer consisting of 46 m*M* sodium phosphate, 18 m*M* sodium borate, and 18 m*M* sodium ascorbate, pH 7.0. Incubations were performed at room temperature for different time intervals and the released amino acids were analyzed directly. The hog kidney aminopeptidase M (Boehringer Mannheim) used in this N-terminal analysis was prepared by dialysis against 50 m*M* sodium phosphate, pH 7.0, at a nominal concentration of 0.09 mg/ml.

Gel electrophoresis and isoelectric focussing

Acetic acid-urea polyacrylamide gel electrophoresis (AU-PAGE) was performed according to Panyim and Chalkley (1969) with 5% polyacrylamide, 5% acetic acid and 8 M urea, pH 2.7 (Rzepecki *et al.*, 1991). Discontinuous sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli (1970), except that various concentrations of DTT were used in the sample treatment buffer. Low molecular mass protein standards ranged from 14.4–97.4 kDa (Bio-Rad, California). Gels were stained for protein in 0.001% Coomassie Blue R-250 in 7.5% acetic acid, 40% methanol.

The DOPA proteins were visualized by staining the gels with the Arnow (1937) method, as described by Waite and Tanzer (1981). We also used a new redox cycling method involving the production of formazan from nitroblue tetrazolium (NBT) in the presence of glycine at pH > 10 (Paz *et al.*, 1991), but at a small cost in sensitivity, did not usually transfer the proteins electrophoretically to nitrocellulose as recommended by the authors. For Arnow assays, SDS gels were first acidified by equilibration in 5% acetic acid; for NBT stain, all gels were washed twice in 50–100 ml of 0.2 *M* sodium borate, pH 8.5, to

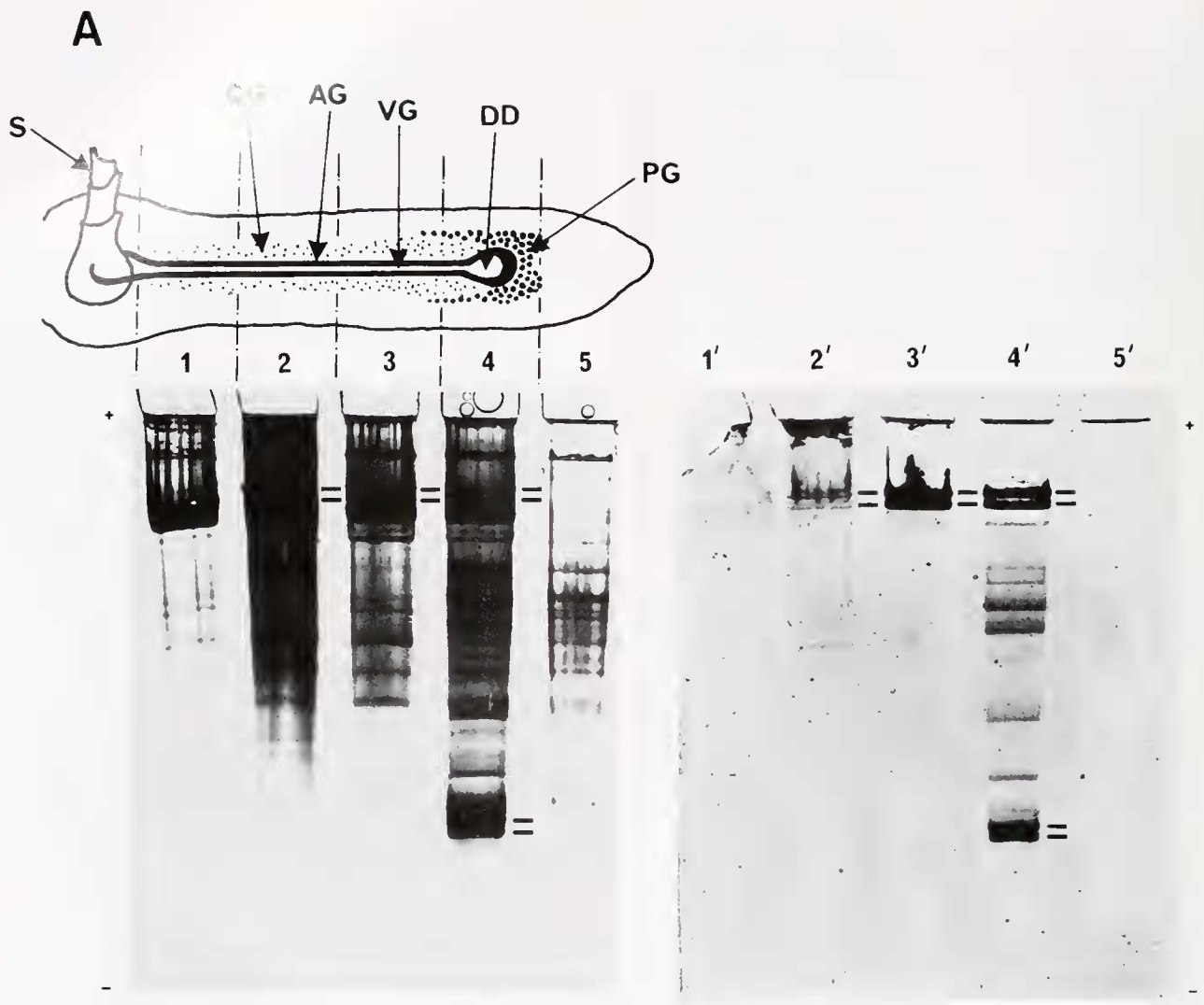

Figure 1. Localization of Mefp-2 in *Mytilus edulis* foot and byssus. (A) AU-PAGE analysis of acid-urea soluble proteins from a serially sectioned mussel foot. Lanes 1–5 correspond to the indicated foot sections and were stained with Coomassie Blue. Lanes 1′–5′ correspond to the same sections run on a parallel gel stained by the NBT redox cycling assay. The drawn parallel bars indicate position of Mefp-1 (upper) and Mefp-2 (lower) proteins. Schematic codes are: S, stem of the byssus; CG, collagen gland; AG, accessory gland; VG, ventral groove; DD, distal depression; and PG, phenol gland. (B) SDS-PAGE (15% polyacrylamide) of proteins extracted from threads and plaques in SDS and neutral buffer. Gels stained as indicated; symbols P, T, M-2 and St refer to: plaque extract, thread extract, pure Mefp-2, and molecular weight standards, respectively; arrows indicate Mefp-2. (C) AU-PAGE analysis of acetic acid-urea soluble proteins from threads and plaques. Gels stained as indicated; symbols are as in (B); arrows indicate Mefp-2. The higher molecular weight aggregates apparent in pure Mefp-2 in (B) and (C) were artefactual results of prolonged storage in water at −20°C and were never found in freshly isolated protein. Arnow tests were not performed in (B) and (C). (D) Dot blots of pure DOPA proteins, and thread and plaque acetic acid-urea soluble proteins, stained as indicated. The amount of total protein blotted is given in the center: symbols are as in (B).

remove the electrophoresis buffers before adding the NBT redox cycling reagents. Use of the NBT redox cycling assay was prefered in this study since, though it is somewhat less specific than the Arnow assay in detecting *o*-diphenols, it is considerably more sensitive, and the stain lasts indefinitely, once the gels have been washed first in alkaline borate and then in 5% acetic acid. Unless specified, however, the presence of DOPA in electrophoretic protein species was confirmed at least once by the Arnow stain before routine use of the NBT assay.

Glycoproteins were detected by the periodic acid-Schiff (PAS) stain (Segrest and Jackson, 1972), with bovine α_1-

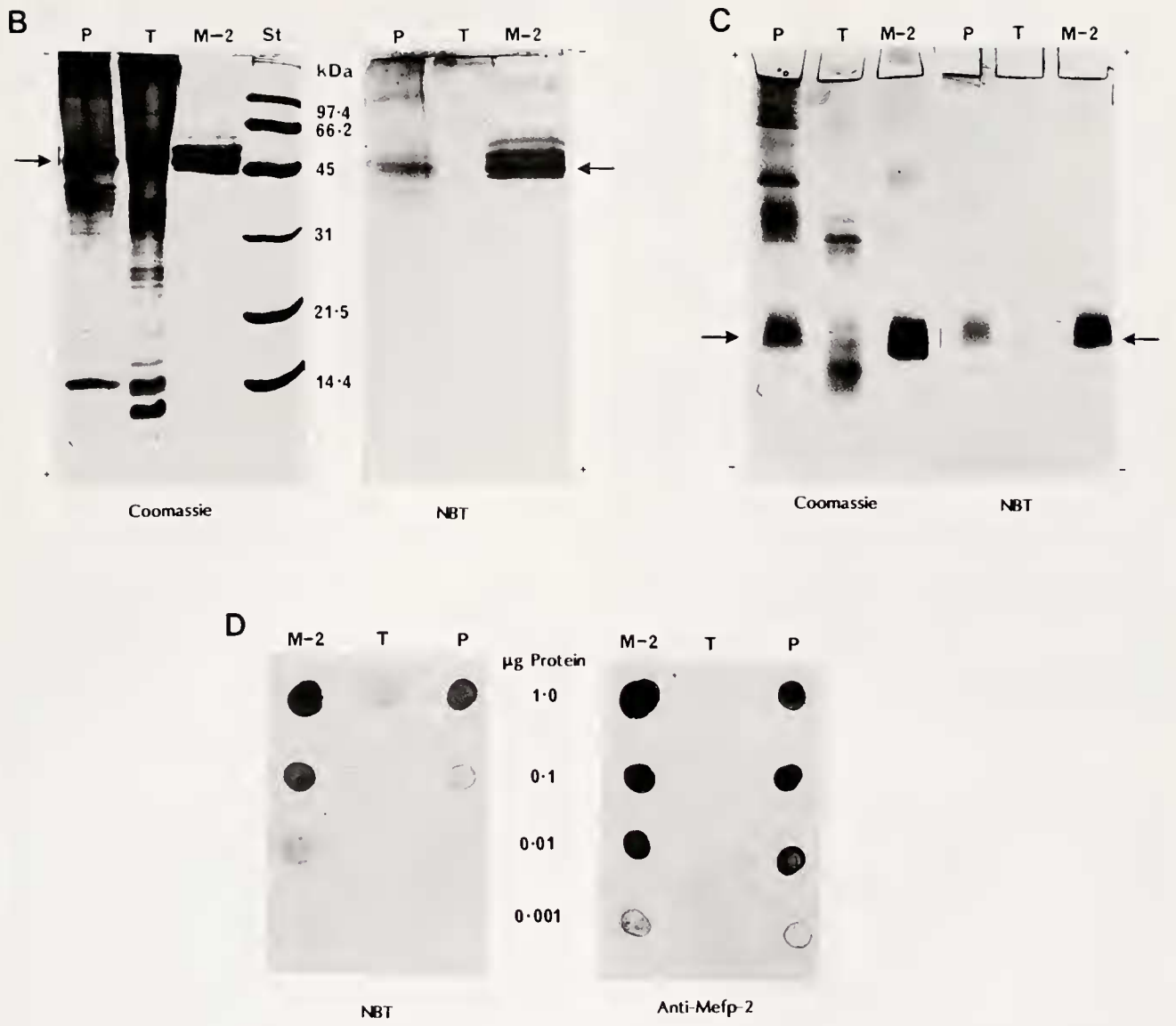

Figure 1. *(Continued)*

acid glycoprotein (Sigma Chemical Co., St. Louis, Missouri) and collagen (type II calf skin) serving as positive controls.

Denaturing isoelectric focussing (IEF) was performed with pH 3–10 ampholytes (FMC Bioproducts, Rockland, Maryland) in 8 *M* urea, 10% glycerol, 6% triton X-100, using 0.02 *M* acetic acid and 1 *N* NaOH as the anolyte and catholyte, respectively (Guilian *et al.*, 1984). Gels were pre-equilibrated (200 V, 15 min; 500 V, 30 min; 800 V, 30 min) before samples (in the above buffer with 15 mg/ml DTT) were applied to the gel under the anolyte. Proteins were focussed for 2–3 h at 1000 V, and a lane was removed for pH gradient determination (by incubation of 0.5 cm gel slices overnight in 1 ml of 0.1% NaCl) before the gel was stained.

Results and Discussion

Localization of Mefp-2 in M. edulis *byssus*

DOPA proteins extracted from serially sectioned mussel feet and electrophoretically resolved by AU-PAGE were detected by the NBT redox cycling assay and with Coomassie Blue (Fig. 1a). Mefp-1 migrated as a doublet identical to the previously characterized 130 kD DOPA protein (Waite, 1983), and it was found in foot sections throughout the length of the ventral groove corresponding to accessory and phenol glands. Mefp-2 was exclusively associated with the distal foot section that encompasses the phenol gland involved in plaque secretion. Several minor NBT-positive proteins of intermediate electrophoretic

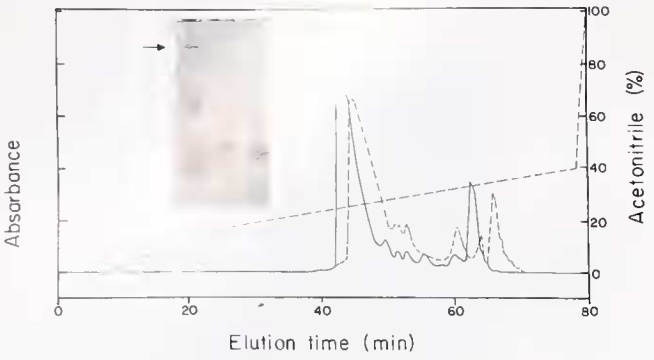

Figure 2. HPLC chromatography profiles of Mefp-2 (———) and Mefp-1 (-----) monitored at 280 nm. Proteins were resolved on a 0.7 × 25 cm semi-preparative C_8 column with an acetonitrile gradient, as indicated by the inclined dashed line. *Inset:* AU-PAGE of purified Mefp-2 at three concentrations, stained with Coomassie Blue. The arrow indicates the position of trace Mefp-1 contaminants.

mobility were also observed in this distal segment, but their origin and nature were not investigated further. To obtain a better separation between phenol and accessory glands, we micro-dissected phenol and accessory glands from thin foot sections and again, Mefp-2 was exclusively detected in the phenol gland, whereas Mefp-1 appeared in both glands. No other byssal glands are known to contain DOPA proteins.

Extraction of byssal plaques in either acetic acid-urea or neutral SDS buffers yielded a polypeptide that co-migrated with Mefp-2 during both SDS- and AU-PAGE and stained positively with NBT (Fig. 1b, c); miscellaneous electrophoretic species that stained with Coomassie Blue, but not NBT, also appeared. Preliminary amino acid analysis following HPLC purification of this NBT-positive plaque polypeptide revealed a composition very similar to authentic Mefp-2 (Diamond and Waite, unpub.). No polypeptide extracted from byssal threads alone exhibited NBT reactivity or electrophoretic behavior akin to Mefp-2, and no polypeptides corresponding to intact Mefp-1 were detected in plaque or thread extracts. Dot blots confirmed that a considerable proportion (*ca.* 10%) of acid-soluble plaque protein was NBT-sensitive, and immunoreactivity assays demonstrated that the extracted material that reacted with anti-Mefp-2 antibodies was specific to plaques (Fig. 1d). Pre-immune serum was unreactive in these assays. These observations are complicated by some cross-reactivity between the as yet unpurified polyclonal anti-Mefp-2 and pure Mefp-1, but the electrophoresis and dot blot results together strongly suggest that, because Mefp-1 appeared to occur in both threads and plaques (Benedict and Waite, 1986a), Mefp-2 is an important component of the plaque.

Purification and composition of Mefp-2

Considerable separation of Mefp-1 and -2 was obtained by chromatography of the proteins on Sephadex G-150 or G-200 in 5% acetic acid (G-200 gave better separation but was not reuseable). Mefp-2, the predominant component of the trailing Sephadex fractions, was rechromatographed on Sephacryl S-300 or S-400, because HPLC alone failed to separate Mefp-2 from residual Mefp-1. Although the difficulty of assessing the relative contributions of either DOPA protein to an unknown mixture preclude presentation of a quantitative purification table, final yields of pure Mefp-2 ranged between 1–5 mg from 30 g wet mussel feet, depending on the season of the year in which the mussels were collected (mussels collected in winter gave better yields than in other seasons). Mefp-2 was resolved by AU-PAGE into two predominant electrophoretic bands (Fig. 2, inset), although minor bands

Table 1

Amino acid composition of Mefp-2 and related proteins[a]

Amino acid, symbol	Mefp-2	CAMC-Mefp-2	Mefp-1[b]	Plaques[c]
3-Hyp	0.0	0.0	42–27	1.6
4-Hyp	0.0	0.0	161–120	28.4
Asp, D or Asn, N	127.0	115.1	12	88.3
Thr, T	42.2	38.8	113	32.6
Ser, S	73.9	65.8	93	66.5
Glu, E or Gln, Q	49.2	43.7	5	50.0
Pro, P	111.9	91.8	41–82	44.0
Gly, G	141.5	132.6	10	219.9
Ala, A	38.6	35.3	81	89.2
Hcys	68.2	1.8	0	20.5
Cys, C	0.0	155.3	—	—
Val, V	43.4	37.3	5	34.0
Met, M	1.9	1.0	2	5.9
Ile, I	9.0	8.2	8	20.3
Leu, L	12.6	11.7	trace	44.6
Dopa, Y* (or ¥)	28.6	25.8	181–110	14.8
Tyr, Y	56.0	51.7	31–73	46.2
Phe, F	9.6	9.0	0	10.0
His, H	8.0	10.1	3	46.0
Lys, K	134.7	124.7	214	51.8
Arg, R	43.7	40.2	3	61.0

[a] Amino acid compositions are given as residues per thousand (RPT) and are averaged from several preparations with a standard deviation of approximately 5%. The Mefp-2 composition was derived as described in Materials and Methods to obtain both DOPA and cystine values. CAMC-Mefp-2 was generally hydrolyzed by the rapid hydrolysis method, and cysteine was quantitated as *S*-carboxymethylcysteine. Ranges of values for certain amino acids in the Mefp-1 composition were derived from successive fractions under the peak of HPLC purified protein, and reflect differing degrees of hydroxylation of Pro and Tyr. Few such differences were apparent in Mefp-2. Cys, Lys and DOPA values are underlined for convenience.

[b] Rzepecki *et al.* (1991).

[c] Benedict and Waite (1986b).

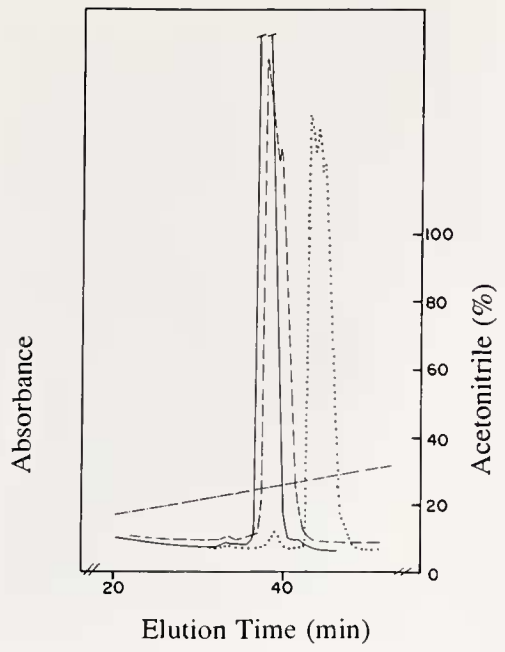

Figure 3. HPLC chromatography profiles of control Mefp-2 (———), and CAMC-Mefp-2 (-----) and COMC-Mefp-2 (· · · · · · · ·), reduced and alkylated as indicated in Materials and Methods, and monitored at 280 nm. Proteins were resolved on a 0.46 × 25 cm analytical C_8 column using the same gradient program as shown in Figure 2 (inclined dashed line). Concentration of control Mefp-2 was half that of the alkylated derivatives, but the detector attenuation was halved.

of similar mobility were often apparent. All reacted positively with the NBT and *o*-diphenol-specific Arnow stains (not shown), consistent with presence of DOPA in the amino acid composition. Mefp-2, resolved by SDS-PAGE (see below), did not stain with the PAS procedure for *cis*-diol sugars, though control proteins in the same gel were PAS-positive, suggesting that Mefp-2, in common with Mefp-1, is not a glycoprotein.

The most dramatic compositional difference between Mefp-1 and -2 (Table I) is the occurrence of cystine exclusively in Mefp-2 and at remarkably high levels (6–7 mol %). A comparison of native Mefp-2 with reduced and *S*-carbamidomethylated protein (CAMC-Mefp-2), together with the exceedingly low levels of *S*-carboxymethylcysteine in the acid hydrolysates of native Mefp-2 that had been treated with iodoacetamide in 4.8 or 6 *M* urea without prior reduction by DTT, shows that most of the *S*-alkylated cysteine (*ca.* 15 mol %) had originally been in the disulphide form. The DOPA content of Mefp-2 (2–3 mol %) was much lower than that of Mefp-1 (11–18%) and, unlike other molluscan DOPA proteins (Rzepecki *et al.*, 1991), the Lys:DOPA ratio far exceeded unity. Both contained proline, but no hydroxylation to 3- and 4-Hyp occurred in Mefp-2. Glycine, rare in Mefp-1, made up about 14 mol % of Mefp-2. Elevated levels (about 5-fold) of Asx and Glx, and lower levels of Ser and Thr, were also found in Mefp-2, in addition to other significant differences.

A comparison of the Mefp-1, Mefp-2, and byssal plaque amino acid compositions allows a crude estimate of the contributions of the two DOPA proteins to the overall plaque composition. Assuming that Mefp-1 is the only source of 3-hydroxyproline in byssus (Waite, 1983), the amount of Mefp-1 in the plaque may be estimated from the plaque 3-Hyp value (Benedict and Waite, 1986b) to be 4–5 mol % of plaque protein, and the proportion of DOPA residues belonging to Mefp-1 can be calculated. On the further assumption that all the remaining DOPA belongs to Mefp-2, a round figure of 25 mol % may be calculated for the plaque content of Mefp-2. At this postulated concentration, Mefp-2 would account for about 90% of plaque cystine and 70% of plaque lysine residues. Although these precise numbers should be taken *cum grano salis* [since (i) 3-Hyp occurs at low levels (<0.2 mol %) in the plaque and is difficult to quantitate; (ii) the content of the readily oxidized DOPA residues is easily underestimated, even in fresh byssus; and (iii) minor proteins containing DOPA or 3-Hyp may occur], they are consistent with the apparent proportion of Mefp-2 extractable from plaque and foot (Fig. 1). These considerations thus establish an upper limit to the content of Mefp-2 in the plaque.

HPLC and electrophoretic characterization of Mefp-2

HPLC and electrophoretic analyses of native, as well as reduced and *S*-alkylated Mefp-2 revealed a challenging complexity in chromatographic and electrophoretic behavior. The alkylated Mefp-2 derivatives eluted more slowly than control protein on HPLC, with the *S*-carboxymethylated protein (COMC-Mefp-2) unexpectedly the slower of the two (Fig. 3). Mefp-2 that had been reduced with DTT, but not alkylated, eluted with a delay comparable to COMC-Mefp-2 (not shown). Denaturing AU-PAGE profiles of fractions under the chromatographic peaks of Figure 3 showed that the two major bands exhibited by native Mefp-2 had been converted, upon reduction and alkylation, into multiplets with four or more components (Fig. 4a). Both COMC- and CAMC-Mefp-2 migrated considerably more slowly than the native polymer, with COMC-Mefp-2 again the slower of the pair. Mefp-2 that had been reduced by DTT, but not alkylated, migrated with intermediate mobility, but upon storage in water at 4°C, partially regained the faster mobility characteristic of native Mefp-2 (not shown). All electrophoretic species stained positively for DOPA with both NBT and Arnow tests (not shown).

Electrophoretic mobility on denaturing AU-PAGE systems depends both on molecular weight and charge density (Hollecker, 1990). Although alkylation of Mefp-2

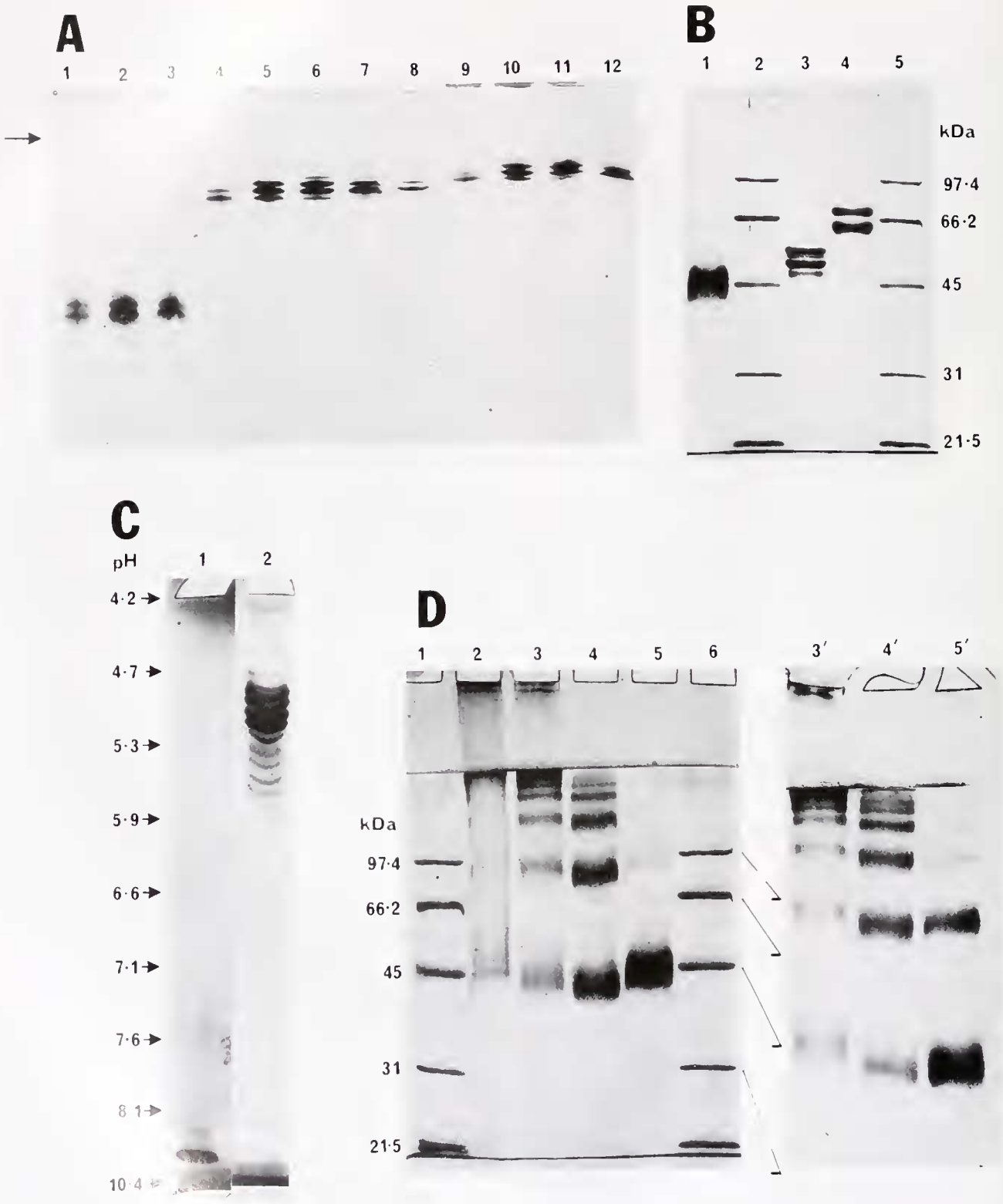

Figure 4. (A) AU-PAGE analysis of sequential HPLC fractions under the chromatographic peaks of Figure 3. Lanes 1–3, native Mefp-2; lanes 4–8, CAMC-Mefp-2; lanes 9–12, COMC-Mefp-2. The arrow indicates the expected position of Mefp-1. Gel stained with Coomassie Blue. (B) SDS-PAGE (12% poly-

would increase the molecular mass of the protein by some 3–3.5 kDa (assuming about 56–62 alkylated cysteines per protein of 42–47 kDa with 360–400 amino acids), this increase alone seems insufficient to account for the reduced mobility under AU-PAGE, so modified charge density and alterations in protein conformation induced by disulphide reduction and alkylation undoubtedly contributed to the electrophoretic differences. The overall similarity between the relative mobilities of CAMC- and COMC-Mefp-2, which have widely different pI values (see below), together with their common difference from native Mefp-2, which has the same pI as CAMC-Mefp-2, argue for the pre-eminence of conformational changes. A transformation from a compact to an extended configuration upon reduction and alkylation is consistent with the observed anomalous electrophoretic and chromatographic migration.

Analysis of Mefp-2 and its alkylated derivatives by SDS-PAGE, following reduction in 50 m*M* DTT, was similarly complex (Fig. 4b). In contrast to Mefp-1, which precipitates in SDS, native Mefp-2 usually exhibited two electrophoretic bands (at low loading concentrations) with apparent molecular masses in the range 42–47 kDa, although minor NBT-positive bands migrating in the same range were frequently present. Electrophoretic migration was, however, somewhat variable within that range and appeared dependent on the protein concentration and degree of reduction. Several electrophoretic bands [which exhibited positive NBT and Arnow reactions (not shown)] were resolved for CAMC- and COMC-Mefp-2, with apparent molecular masses in the ranges 48–54 kDa and 63–71 kDa, respectively. The decreased mobility of CAMC-Mefp-2 was almost, yet not quite, explained by the molecular mass increment (3–3.5 kDa, see above) resulting from alkylation, but COMC-Mefp-2 migrated considerably more slowly. *S*-Alkylation with iodoacetate is well known to induce electrophoretic anomalies (Lane, 1978).

Isoelectric focussing (IEF) studies in highly reducing and denaturing media (Fig. 4c) showed that the apparent pI of native Mefp-2 lay between 9 and 10 pH units (*i.e.*, largely beyond the resolving power of available ampholytes), similar to that for Mefp-1. CAMC-Mefp-2, which theoretically has the same net charge as native Mefp-2, migrated identically (not shown). COMC-Mefp-2, with a greatly decreased net positive charge owing to the additional carboxyl moieties, migrated as a family of at least ten distinct members, with pI values in the range 4.7–5.9. Analogous electrophoretic heterogeneity occurs in the DOPA proteins of *Fasciola hepatica* (Waite and Rice-Ficht, 1989).

The variable migration of native Mefp-2 during SDS-PAGE caused us to investigate further the effects of DTT on electrophoretic behavior in this system. In the absence of DTT, Mefp-2 could barely be induced to enter even the stacking gel of discontinuous SDS-PAGE systems (Fig. 4d). As the DTT concentration was increased to 50 m*M*, progressively more Mefp-2 entered the resolving gel and migrated in a ladder-like pattern consistent with the presence of higher molecular weight aggregates of 2, 3, 4 or more polypeptide chains. At the highest DTT concentrations, most of the Mefp-2 migrated as an apparent monomer (*cf.* Fig. 4b). The obvious explanation for this observation, that the oligomeric Mefp-2 resulted from intermolecular disulphide bonds, does not seem tenable, since significant high molecular weight aggregates were rarely observed upon resolution of the same protein preparations by denaturing AU-PAGE (although they sometimes occurred on prolonged storage when frozen in water—*cf.* also Figs. 1b, c and 2, with Fig. 4a, d). Apparently, then, conditions exist for the aggregation of native, but not reduced, Mefp-2 in solution, although the precise contribution, if any, of SDS to this phenomenon remains to be determined. This is consistent with the precipitation frequently noted in concentrated neutral solutions of Mefp-2, and with preliminary X-ray scattering data (Trumbore, pers. comm.).

Primary structure of Mefp-2

Despite the apparent electrophoretic heterogeneity of Mefp-2, N-terminal sequencing of pure protein revealed a single detectable N-terminus, H_2N-Thr-Ser-Pro-Xaa-Yaa-Dop-Asp-Asp-Asp-Glu . . . , where Xaa and Yaa

acrylamide) analysis of Mefp-2 and derivatives. Lane 1, native Mefp-2; lane 3, CAMC-Mefp-2; lane 4, COMC-Mefp-2; lanes 2 and 5, molecular weight standards. Gel stained with Coomassie Blue. (C) IEF analysis of Mefp-2. Lane 1, native Mefp-2; lane 2, COMC-Mefp-2; CAMC-Mefp-2 migrated identically to native Mefp-2 and is not shown. Proteins were focussed on separate gels in parallel, and the gels were stained with the NBT redox cycling assay. The pH gradient is indicated on the left. (D) Effects of DTT on native Mefp-2 migration during SDS-PAGE (12% polyacrylamide). Lanes 1 and 6 are molecular weight standards. DTT concentrations: lane 2, 0 m*M*; lane 3, 0.5 m*M*; lane 4, 5 m*M*; lane 5, 50 m*M*. The gel was stained with Coomassie Blue. Lanes 3′–5′ are from a gel run in parallel, but stained with the NBT redox cycling assay, and correspond to lanes 3–5 in the gel on the left.

were predominantly P [illegible] Arg, though Tyr, Gly, Thr and Gln were also [illegible] Table II). Treatment of Mefp-2 with aminopept[illegible] released low levels of Thr, Ser, Pro, DOPA [illegible] consistent with the sequenator analysis. [illegible] ical combination of variability and constancy at the N-terminus might be explicable in terms of multiple genes or alternate mRNA splicing mechanisms (*cf.* Bobek *et al.*, 1988; Pihlajaniemi and Tamminen, 1990), but analysis of Mefp-2 from an individual organism has not yet been possible.

Native Mefp-2 (unlike Mefp-1) was relatively resistant to a variety of proteases, including trypsin, chymotrypsin, pepsin, *Staphylococcus aureus* V8, and collagenase, even at unusually high protease:Mefp-2 ratios (up to 1:1) in the presence of urea, although increases in electrophoretic mobility and heterogeneity were often apparent upon electrophoresis (see Fig. 5a, inset, for AU-PAGE of trypsin-treated Mefp-2). SDS-PAGE in the presence of DTT showed that, upon trypsin treatment, the apparent molecular mass decreased from about 45 kDa to 30–35 kDa with some increase in heterogeneity (not shown). HPLC of trypsin-treated native Mefp-2 resolved some minor peptides (Fig. 5a), but reduction and alkylation of the residual disulphide bonded protein did not result in the recovery of any new peptides upon rechromatography, confirming that little internal nicking had occurred. Trypsin digestion of both COMC- and CAMC-Mefp-2 converted both alkylated derivatives to their component peptides (Fig. 5). These results suggest that trypsin trimmed native Mefp-2 at the N- or C-terminus, or both, leaving the residual disulphide bonded core protein largely intact.

Peptides obtained by HPLC chromatography of trypsin-treated native and alkylated Mefp-2 fell generally into three major classes, all containing either tyrosine or DOPA (Figs. 6, 7). All three types were isolated from alkylated Mefp-2, but only Type II peptides were identified in digests of native Mefp-2. Owing to the chromatographic complexity of the peptide mixture, the relative content of the three classes was rather difficult to determine with any precision. Most fractions contained two major peptides (usually in unequal proportions), and sequences for peptides in various HPLC fractions were assigned by consideration of relative yields of amino acids in successive sequenator cycles. Fortunately, sequence consistencies became readily apparent, and a consideration of Figures 5–7, assuming that UV absorbance at 230 nm broadly reflects peptide concentration, suggests that the majority (at least 80%) of Mefp-2 can probably be accounted for by the identified motifs.

Type I peptides. This type of motif appeared to predominate in Mefp-2, with identified peptides accounting for perhaps 40–50% of the 230 nm absorbance of Figure 5b, although no individual peptide contributed more than 5–10%. It was highly enriched in Cys, Gly, and Pro, and was basic due to the excess of Lys and Arg over Glu and Asp. Significantly, we found no evidence that the Tyr residues were hydroxylated to DOPA in this type of motif. For convenience, we have subdivided the Type I motif into six submotifs (Fig. 6), which combine variously to

Table II

N-terminal analysis of Mefp-2[a]

	Sequenator analysis Amino acid (yield, pMoles × 10^3)		Aminopeptidase analysis
Cycle #	Fraction 5	Fraction 6	Amino acid (RPT)
1	T (2.0)	T (1.5)	S (225, 219)
2	S (1.8)	S (1.4)	T (185, 178)
3	P (2.2)	P (1.8)	P (171, 187)
4	R (2.1)	P (1.1) R (0.6) Y (0.5) G (0.5)	R (70, 88)
5	P (1.0) T (0.5) Y (0.5) Q (0.4)	P (1.2) Q (0.9)	Y* (64, 48)
6	Y* (1.7)	Y* (1.4)	D/N (32, 25)
7	D (1.5)	D (1.6)	[G (93, 87)]
8	D (1.6)	D (2.1)	
9	D (2.3) G (1.2)	D (2.1)	
10	several amino acids detected in significant quantities (D, E, G, K, Y)		

[a] N-Terminal sequences and aminopeptidase M analyses were obtained as described in Materials and Methods. Sequences were derived from two adjacent fractions under the peak of an HPLC elution profile of Mefp-2, similar to that of Figure 2. In the aminopeptidase analysis, amino acid compositions from two individual experiments (with aminopeptidase incubation times of 2.5 and 16 h), are reported as residues per thousand (RPT) for the most abundant amino acids, the remainder ranging between 0–30 RPT. The yields were low, perhaps due to protein aggregation, adsorption in neutral buffer, or both; thus the glycine detected probably reflected background contamination of the vessels or protein preparations. One letter amino acid codes are given in Table I.

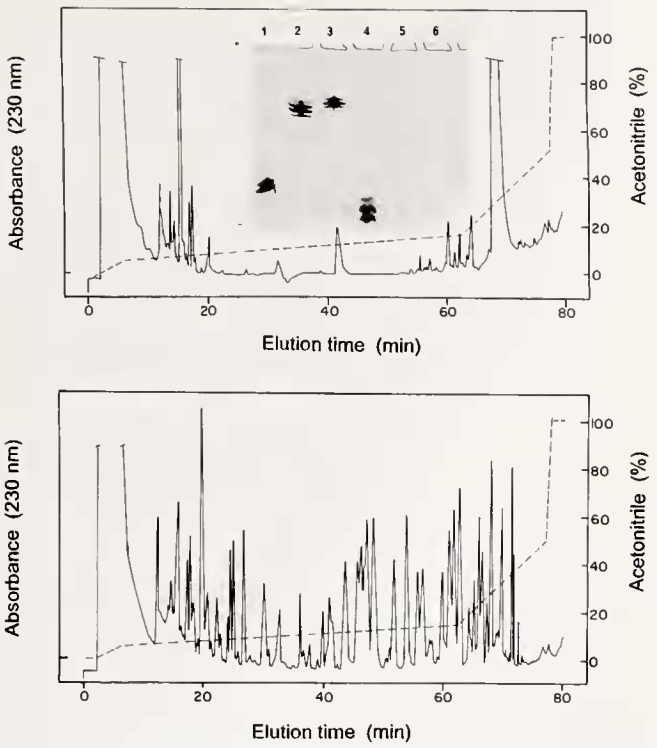

Figure 5. HPLC chromatography profiles of Mefp-2 and derivatives following proteolytic digestion with trypsin. Peptides were resolved on a 0.46 × 25 cm C_{18} column using acetonitrile gradients as indicated by the dashed lines. Fractions of 1 ml were collected at a flow rate of 1 ml/min. (A) Peptides and undigested core protein (fractions 68–72) derived from native Mefp-2. *Inset:* AU-PAGE analysis of native Mefp-2, CAMC-Mefp-2 and COMC-Mefp-2 before (lanes 1–3, respectively), and after (lanes 4–6, respectively) proteolysis with a trypsin:protein weight ratio of 1:50, or 1:25 in the case of native Mefp-2. Equivalent concentrations of protein were run in each lane, and the gel was stained with Coomassie Blue. (B) Peptides derived from COMC-Mefp-2. The peptide map derived from CAMC-Mefp-2 was qualitatively similar to that shown in (B).

Cystine Enriched Tryptic Peptides

Oligopeptide Submotifs

Peptide # COMC	CAMC	Sequence
		Consensus (*Type I*): [CY]CV GGYSG PTC QENAC KPNPC [X]K T L GV V A
		Type I peptides
38	32b	*C APLGK*
50b		VNVC KPN...
64b	53b	*YIC APNPC K*
57a	46b	*YVC APNPC K*
48a		C GENVC KPNPC QNK V A
49a		C QVHAC KPNPC N{N/T}K
52a		C GVNAC NPNPC K
62a	51b	*GGYSG PTC GENVC APNPC QCK*
63a	52b	*GGYSG PTC GVNAC KPNPC QNR* F
63b	51a	*CS GGYSG PTC NVNAC KPNP- NK*
64a	53a	*CV GGYSG PTC QENAC KPNPC K*
66a		CTCV GGYSG PTC QENAC KPNPC SN...
69a		CL GGYIG PTC -ED— KPNPC N...
68		CL GGYKG PTC -ED— KPNPC NIK V NV
	54b	CYCV GGYSG FQ...
71		CYCT GGYCG LT...
58	54a	*CYHD NSDDG PTC QENAC KPNPC N...*
55b	52b	*CYHD K*
31	27/28	*CYHD GK*
34	32a	*GGYSG P*(K)
53	48	*[S]GYYG PEC ...* Q
49b		GTGYVGYIG R
55a		GSSYN- --C ---IC K
		Others
73a		[CSADK]FGDYTCDFRPGYFGPE{C/Q}...
69b		GGYGKFNDYLE...

Figure 6. Cystine-enriched tryptic peptide sequences derived from alkylated Mefp-2. Numbers under the COMC heading correspond to fractions obtained from the HPLC chromatogram depicted in Figure 5b (*i.e.*, peptides from COMC-Mefp-2); those under CAMC correspond to HPLC fractions of peptides from CAMC-Mefp-2 (not shown). Letters after the numbers indicate whether the peptide was a major (a) or minor (b) component of the HPLC fraction. In cases of ambiguity, where more than one peptide was found in the fraction, sequences were assigned by consideration of relative PTC-amino acid yields in successive sequenator cycles. Sequences in *italics* are those obtained from both COMC- and CAMC-Mefp-2 preparations. DOPA and alkylated cysteine residues are emphasized, and sequences are spaced as submotifs for convenience only. Residues in parentheses () were inferred from composition data, or were difficult to assign owing to low yields or ambiguous results in a sequenator cycle. Residues in brackets [ ] are optional sequence extensions in peptide variants. Residues separated in braces {/} are alternative amino acids found in that position of the sequence; in other cases, alternative amino acids are shown below the main peptide sequence. Ellipses (. . .) indicate probable continuation of a peptide beyond the last residue clearly resolved on the sequenator; dashes (– – –) indicate gaps introduced into the sequence to optimize putative alignment. The symbol X in the consensus sequence indicates one or two optional small polar amino acids.

form both complete and truncated motifs. Tyrosine typically occurs in submotif #2, GGYSG, in the company of small residues. This propinquity of Tyr and small residues is common in tyrosine- or DOPA-containing structural proteins from species in various phyla (*e.g.*, Johnson *et al.*, 1987; Waite and Rice-Ficht, 1987, 1989; Bobek *et al.*, 1988; Mehrel *et al.*, 1990; Aggeli *et al.*, 1991; Rzepecki *et al.*, 1991); it has been attributed in part to the hypothetical steric requirements of 3-tyrosyl hydroxylases (Rzepecki *et al.*, 1992), although this feature is clearly insufficient for tyrosine hydroxylation in Mefp-2. Tyrosine also occurred in submotif #1 with vicinal Cys residues; this relationship has never been observed in a protein where tyrosine is known to be hydroxylated to DOPA, but occurs infrequently in other glycine, cystine, and tyrosine enriched structural proteins, although replacement of Tyr by Cys in $(G)_nY/C$ sequences is more usual (*e.g.*, Bobek *et al.*, 1988; Mehrel *et al.*, 1990; Aggeli *et al.*, 1991).

Submotifs #3 (PTC) and #5 (KPNPC) are enriched in Cys and Pro and are, in addition to the GGYSG submotif, the most conserved [illegible]x. Submotif #4 is relatively variable, with [illegible]s of Gly for Gln, Val for Glu, and Ala for V[illegible] the terminal Cys residue appears invariant. The [illegible]-terminal submotif (#6) consists of 0–2 small polar residues (Asn, Gln, etc.) with a C-terminal Lys or Arg. A search of the SWISS-PROT Protein Sequence Data Bank [release 19 (August 1991) with 21,795 entries] failed to identify any remarkable homologies between the Type I motif and any other known sequence. [Note, however, the conceptual similarity between Type I motifs and conotoxin sequences. Conotoxins, isolated from the venomous marine cone snails (*Conus*), are small (10–30 amino acid) multiply disulphide bonded peptides, with 2–4 highly variable loops of 1–6 amino acids between adjacent Cys residues, and are potent antagonists of various neuromuscular receptors (Olivera *et al.*, 1990)]. The Cys-rich nature of Type I motifs, as well as their relative abundance, suggests that they are derived from the core of Mefp-2, which is resistant to trypsin digestion in the non-alkylated state. If they are thus concentrated, then such an arrangement would result in a quasi-periodic distribution of Cys and Tyr residues over much of the protein.

Type II peptides. These motifs were enriched in Pro and Ser, and were highly acidic, containing almost palindromic runs of Asp and Glu bracketed by DOPA residues (Fig. 7). The first DOPA was often found at the C-terminus of a PPSY* (Y* = DOPA) sequence, analogous to a submotif of the Mefp-1 tandem repeat (AKP*SY*P*P*TY*K where P* = hydroxyproline; Waite, 1983). The second DOPA had Asp as its N-terminal partner with, frequently, Thr at its C-terminus; similar arrangements occur in the high molecular weight DOPA proteins from other molluscan species, but are foreign to Mefp-1 (Rzepecki *et al.*, 1991). No Cys or Gly residues were found, thus sealing the distinction between Type I and Type II sequences. The Mefp-2 N-terminus, TSPxxY*DDD(E) . . . (Table II and Fig. 7), is a Type II sequence, and was the only N-terminus detected (see above). The C-terminal sequence (K)SPPSY*NDDDEY* was identified in several independent peptide preparations by the close correspondence between the observed amino acid composi[illegible] which lacked Lys or Arg (in trypsin derived prepar[illegible]) but had equimolar DOPA, Ser, and Pro contents, a[illegible] expected composition calculated from the sequen[illegible] ultimate DOPA residue was inferred largely fro[illegible]sition data, since sequenator yields of C-terminal [illegible]ere greatly reduced, though still marginally eviden[illegible] some cases. No other peptides were reliably identified a[illegible]-terminal. Thus the DOPA rich N- and C-terminal sequences of Mefp-2 are probably structurally similar and, in contrast to the basic Type I

Dopa Enriched Tryptic and Chymotryptic Peptides

COMC	Peptide # CAMC	Native	Sequences	Comments
			Type II peptides	
		†Untreated	TSPxx**Y**DDD (E) . . .	N-Terminus
66b			TSPDP**Y**DDDEDD**Y**TPPV (R)	
62b			T----**Y**DDDEDD**Y**TPPV (R) K	
		†Trypsin	TSP----DDEDD?TPPV (K) D P DE Q	Dopa prominent in composition but not detected in sequence.
		16	TSP (R) . . ?	
		†Chymotrypsin	KSPPS**Y**NDDDE (**Y**)	C-Terminus?
		†Trypsin	SPPS**Y**NDDDE (**Y**)	C-Terminus?
45a		43/44	*SPPS**Y**NDDDE (**Y**)*	C-Terminus?
		†Chymotrypsin	KSPPSNNDD . . .	
			Type III peptides	
52b			A**Y**--NP**C**FKR	
47a			A**Y**KPNP**C**ASR[PGYR]	
48a*			A**Y**KPNP**C**ASG-P**C**KK **Y** N . . .	
49*			A**Y**KPNP**C**ASS-P**C** (K)	
58*			A**Y**KPNP**C**VVSKP**C** (**C**) . . .	
			Others	
70*			**Y**FP**CC**PS**Y**NG . . .	
48b			NT**C**I**C**G**Y**NGSGG**Y**R	
50a	46a		*Y**N**GV**C**KP{N/S}GGS**Y**K*	
47b			**Y**NGV . . .	
48b*			**Y** SNGV**C**K (F) SG . . .	
			Y	
19		†Trypsin	G**Y**SGR **Y** K	
62*			**Y**?PVNP**C**LKK	
		70	**Y**SPVN . . . **Y**	Core Protein N-Terminus??
68*			GG**YY**G**Y**N**C**N . . .	

Figure 7. DOPA-enriched tryptic peptides from native Mefp-2 and alkylated derivatives. Numbers under the Native heading correspond to fractions from Figure 5a, those under COMC and CAMC headings are as in Figure 6. Numbers marked with an asterisk (*) correspond to peptides isolated by phenyl boronate affinity chromatography prior to HPLC purification (not shown). The symbol (†) indicates data from earlier preliminary experiments using Mefp-2 prepared from different batches of mussels. "Untreated" Mefp-2 refers to the N-terminal analysis. Peptides from native Mefp-2 specifically labeled as "trypsin" derived were obtained by digestion of Mefp-2 in 100 m*M* Tris, 50 m*M* ascorbate buffer, pH 8.0 for 24 h at room temperature with a trypsin:protein weight ratio of 1:10, then isolated by HPLC on a 0.46 × 25 cm C_8 column with the same elution gradient as used for protein isolation. Peptides were detected at 280 nm. Peptides specifically labeled as "chymotrypsin" derived were obtained by digestion of Mefp-2 in 150 m*M* Tris, 150 m*M* ascorbate buffer, pH 7.5, at room temperature for 6 h with a chymotrypsin:protein ratio of 1:1, then resolved by HPLC, following lyophilization and resuspension in 5% acetic acid, on a 0.46 × 25 cm Vydac C_{18} column with a 0–25% gradient of acetonitrile in water, incorporating 0.1% trifluoracetic acid. Because few peptides were obtained by either of these protocols, the HPLC profiles have been omitted. Other symbols and protocols are as described in Figure 6, except Y which is a symbol for DOPA to facilitate peptide alignment.

motifs and the DOPA-containing motifs from other DOPA proteins, are highly acidic.

Type III peptides. This last major type contained both DOPA and Cys residues, in addition to Pro, Lys, and Ala. The DOPA residue was found in the sequence AY*KPNPC . . . , where the KPNPC segment is identical to submotif #5 of the Type I motif. Although there is some compositional similarity to the Mefp-1 consensus motif, the only sequence identity is in the dipeptide Y*K, which is a common feature of many molluscan DOPA proteins (Rzepecki *et al.*, 1991). The distribution of Type III peptides within the Mefp-2 sequence is unknown.

Minor peptide sequences were apparent, including some unusual vicinal YY and Y*Y dipeptides previously found in DOPA proteins from the mussels *Modiolus modiolus* and *Trichomya hirsuta* (Rzepecki *et al.*, 1991), and certain tunicates (C. Hawkins, pers. comm.). Although variation in Mefp-2 is considerable, the domination of the heterogeneous Mefp-2 family by a limited number of motif variants indicates that it is tandemly repetitive. However, unlike most DOPA proteins, the motif degeneracy is quite marked, and the sequences often diverge markedly from those expected on the basis of previously analyzed DOPA proteins (Rzepecki *et al.*, 1991; Rzepecki and Waite, 1991). Moreover, Mefp-2 exhibits no sequence homologies with any previously characterized cysteine- or cystine-containing proteins, including disulphide bonded collagens (Dublet and Rest, 1991), high sulphur keratins (Fraser *et al.*, 1972), zinc-binding proteins (DuBois *et al.*, 1990; Opipari *et al.*, 1991), metallothioneins (Lerch, 1980; Nemer *et al.*, 1985), keratinocyte loricrins (Mehrel *et al.*, 1990), proteoglycans such as perlecan (Noonan *et al.*, 1991), and adhesive laminins of the neuromuscular junction (Hunter *et al.*, 1989). Mefp-2 thus appears to be a novel class of structural protein, and any short sequence similarities to other classes may be stochastic or result from common structural features related to function.

Conclusions

The blue mussel, *Mytilus edulis*, incorporates at least two distinct DOPA proteins into its byssus. The first of these to be characterized (Waite, 1983), Mefp-1, forms a protective varnish at the interface between byssus and environment, and may also be the adhesive agent at the interface between plaque and substratum (Benedict and Waite, 1986a). The second major DOPA protein of the blue mussel, Mefp-2, appears to be a structural component exclusively of the plaque, contributing up to 25% of plaque protein. (About 90–95% of the dry weight of the *M. edulis* plaque is proteinaceous; Diamond and Waite, unpublished.) Mefp-2 seems to be a tandemly repetitive, multidomain protein, with short, acidic, DOPA-containing N- and C-terminal regions, and a large central domain constrained by quasi-periodic internal disulphide bridges to a compact conformation resistant to proteolytic degradation. The peptide motifs of Mefp-2 are quite unlike those of any other known structural proteins. Speculations about secondary structure are premature, because in tandemly repetitive proteins, the correct order of short peptide motifs over an entire protein cannot be deduced by standard peptide mapping techniques, and it is debatable whether all members of the heterogeneous Mefp-2 family share identical subsets of motifs. Nevertheless, the sequence of the Type I motif (Fig. 6) of Mefp-2 contains proline and/or glycine rich segments that alternate with sequences containing none of these β-turn associated residues, and it is therefore consistent with conformational models, such as the β-meander postulated for certain glycine and tyrosine rich *Schistosoma* eggshell proteins (Rodrigues *et al.*, 1989).

The composition of Mefp-2, incorporating cystine and DOPA, implies some role involving the stabilization of the plaque matrix by covalent disulphide and quinone-derived cross-links. *In vitro*, Mefp-2 can form oligomeric aggregates that might, in time, be stabilized by rearrangement of disulphide bonds to form inter-molecular cross-links. If Mefp-2 indeed constitutes 25% of plaque protein, then it would incorporate 90% of the plaque cystine residues and thus would be virtually its own sole potential disulphide cross-link partner. The DOPA residues might then serve to cross-link the resulting disulphide-linked Mefp-2 homopolymer to other protein components of the plaque, although the apparent extractability of Mefp-2 from day-old plaques suggests that, if such cross-links are real, they form slowly. Much of the internal volume of the blue mussel plaque is occupied by a microscopic foam which does not bind anti-Mefp-1 polyclonal antibodies strongly (Benedict and Waite, 1986a). Disulphide-rich proteins (Kitabatake and Doi, 1987; Okumura *et al.*, 1989), and rigid proteins in general (Graham and Phillips, 1976; Halling, 1981), are known to stabilize foams, and we had previously noted that Mefp-1 should be a good foam stabilizer by virtue of its cross-linking potential (Rzepecki and Waite, 1991). Mefp-2 also appears suited for such a role, although until it can be precisely localized within the plaque matrix, its function will remain conjectural. [Preliminary attempts to extract acid-soluble proteins analogous to Mefp-2 from the foot of the ribbed mussel *Geukensia demissa*, which secretes a high molecular weight DOPA protein (Gdfp) analogous to Mefp-1, but lacks a distinctive foam structure in its plaque (Waite *et al.*, 1989), have failed to detect significant quantities of any putative Gdfp-2 (Rzepecki, unpub. data); a correlation of habitat and plaque ultrastructure with DOPA protein variants in other mussel species should be informative.]

Many questions, however, remain unresolved. The apparent localization of DOPA to the phenol gland is somewhat equivocal. The phenol gland is an apparent site of Mefp-1 synthesis and storage (Waite, 1983; Benedict and Waite, 1986b). The true nature of the phenol gland secretions is thus puzzling—do they contain one or both DOPA proteins? The sequence differences between DOPA-containing motifs in Mefp-1 and Mefp-2 raise questions about the specificity and number of the putative tyrosyl hydroxylases responsible for conversion of peptidyl tyrosine to DOPA, and also about the nature of the catecholoxidase responsible for peptidyl quinone production (although curiously, no catecholoxidase activity has yet been detected in the phenol gland). The reproducibility of specific disulphide bonds, and thus the degree of conservation of secondary and tertiary structure, is unknown. Finally, although we still require a full representative sequence from the Mefp-2 family, data obtained from cDNA or genomic sequences must be interpreted with caution, because the Mefp-2 family of proteins may prove to be a multigene family and may also be subject to alternative splicing of mRNA. The association of such phenomena with many repetitive structural proteins is well documented (*e.g.,* Bobek *et al.,* 1988; Pihlajaniemi and Tamminen, 1990).

Acknowledgments

We thank Dr. Mark Trumbore for helpful discussions and the computer search of the SWISS-PROT Protein Sequence Data Bank, and Tom Diamond for sharing his studies on byssal plaque protein compositions. This work was funded by a grant from the Office of Naval Research.

Literature Cited

Aggeli, A., S. J. Hamodrakas, K. Koitopoulou, and M. Konsolaki. 1991. Tandemly repeating peptide motifs and their secondary structure in *Ceratitis capitata* eggshell proteins Ccs36 and Ccs38. *Int. J. Biol. Macromol.* **13:** 307–315.

Arnow, L. 1937. Colorimetric determination of components of 3,4-dihydroxyphenyl-L-alanine/tyrosine mixtures. *J. Biol. Chem.* **118:** 531–537.

Benedict, C. V., and J. H. Waite. 1986a. Location and analysis of byssal structural proteins of *Mytilus edulis*. *J. Morphol.* **189:** 171–18[illegible]

Benedict, C. V., and J. H. Waite. 1986b. Composition and ultrastructure of the [illegible] of *Mytilus edulis*. *J. Morphol.* **189:** 261–270.

Blake, M. S., K. H. Johnston, G. J. Russel-Jones, and E. C. Gotschlich. 1984. A rapid, sensitive method for detection of alkaline phosphatase-conjugated anti-body on Western blots. *Anal. Biochem.* **136:** 175–176.

Bobek, L. A., D. M. Re[illegible]nd P. T. LoVerde. 1988. Small gene family encoding an egg[illegible] orion) protein of the human parasite *Schistosoma mansoni*. *Mol. Cell. Biol.* **8:** 3008–3016.

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* **72:** 248–254.

Brown, C. H. 1952. Some structural proteins of *Mytilus edulis*. *Q. J. Microsc. Sci.* **93:** 487–502.

Dublet, B., and M. van der Rest. 1991. Type XIV collagen, a new homotrimeric molecule extracted from fetal bovine skin and tendon with a triple helical disulfide-bonded domain homologous to type IX and type XII collagens. *J. Biol. Chem.* **266:** 6853–6858.

DuBois, R. N., M. W. McLane, K. Ryder, L. F. Lau, and D. Nathans. 1990. A growth factor-inducible nuclear protein with a novel cysteine/histidine repetitive sequence. *J. Biol. Chem.* **265:** 19185–19191.

Filpula, D. R., S.-M. Lee, R. P. Link, S. L. Strausberg, R. L. Strausberg. 1990. Structural and functional repetition in a marine mussel adhesive protein. *Biotechnol. Prog.* **6:** 171–177.

Fraser, R. D. B., T. P. MacRae, G. E. Rogers. 1972. *Keratins: Their Composition, Structure and Biosynthesis*, C. C. Thomas, Springfield, IL.

Graham, D. E., and M. C. Phillips. 1976. The conformation of proteins at the air-water interface and their role in stabilizing foams. Pp. 237–253 in *Foams*, Akers, R. J., ed. Academic Press, London.

Guilian, G. G., R. L. Moss, and M. Greaser. 1984. Analytical isoelectric focussing using a high-voltage vertical slab polyacrylamide gel system. *Anal. Biochem.* **142:** 421–436.

Halling, P. J. 1981. Protein-stabilized foams and emulsions. *CRC Crit. Rev. Food Sci. Nutr.* **15:** 155–203.

Hollecker, M. 1990. Counting integral numbers of residues by chemical modification. Pp. 145–153 in *Protein Structure*, T. E. Creighton, ed. IRL Press, Oxford.

Hunter, D. D., V. Shah, J. P. Merlie, and J. R. Sanes. 1989. A laminin-like adhesive protein concentrated in the synaptic cleft of the neuromuscular junction. *Nature* **338:** 229–234.

Johnson, K. S., D. W. Taylor, and J. S. Cordingley. 1987. Possible eggshell protein gene from *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **22:** 89–100.

Kitabatake, N., and E. Doi. 1987. Conformational change of hen egg ovalbumin during foam formation detected by 5,5′-dithiobis(2-nitrobenzoic acid). *J. Agric. Food Chem.* **35:** 953–957.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680–685.

Lane, L. C. 1978. A simple method for stabilizing protein-sulfhydryl groups during SDS-gel electrophoresis. *Anal. Biochem.* **86:** 655–664.

Lerch, K. 1980. Copper metallothionein, a copper-binding protein from *Neurospora crassa*. *Nature* **284:** 368–370.

Mehrel, T., D. Hohl, J. A. Rothnagel, D. B. Longley, C. Cheng, U. Lichti, M. E. Bisher, A. C. Steven, P. M. Steinert, S. H. Yuspa, and D. Roop. 1990. Identification of a major keratinocyte cell envelope protein, Loricrin. *Cell* **61:** 1103–1112.

Nemer, M., D. G. Wilkinson, E. C. Travaglini, E. J. Sternberg, and T. R. Butt. 1985. Sea urchin metallothionein sequence: key to an evolutionary diversity. *Proc. Natl. Acad. Sci. USA* **82:** 4992–4994.

Noonan, D. M., A. Fullet, P. Valente, S. Cai, E. Horigan, M. Sasaki, Y. Yamada, and J. R. Hassell. 1991. The complete sequence of Perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein-receptor, and the neural cell adhesion molecule. *J. Biol. Chem.* **266:** 22939–22947.

Okumura, K., Y. Miyaki, H. Taguchi, and Y. Shimabayashi. 1989. Enhanced stability of protein foam due to disulfide bond formation just after foaming. *Agric. Biol. Chem.* **53:** 2029–2030.

Olivera, B. M., J. Rivier, C. Clark, C. A. Ramilo, G. P. Corpuz, F. C. Abogadie, E. E. Mena, S. R. Woodward, D. R. Hillyard, and L. J. Cruz. 1990. Diversity of *Conus* neuropeptides. *Science* **249:** 257–263.

Opipari, A. W., M. S. Boguski, V. M. Dixit. 1991. The A20 cDNA induced by tumor necrosis factor α encodes a novel type of zinc finger protein. *J. Biol. Chem.* **265:** 14705–14708.

Panyim, S., and G. R. Chalkley. 1969. High resolution acrylamide gel electrophoresis of histones. *Arch. Biochem. Biophys.* **130:** 337–346.

Pardo, J., E. Gutierrez, C. Saez, M. Brito, and L. O. Burzio. 1990. Purification of adhesive proteins from mussels. *Prot. Express. Purif.* **1:** 147–150.

Paz, M. A., R. Flückiger, A. Boak, H. M. Kagan, and P. M. Gallop. 1991. Specific detection of quinoproteins by redox-cycling staining. *J. Biol. Chem.* **266:** 689–692.

Pihlajaniemi, T., and M. Tamminen. 1990. The a1 chain of type XIII collagen consists of three collagenous and four noncollagenous domains, and its primary transcript undergoes complex alternative splicing. *J. Biol. Chem.* **265:** 16922–16928.

Pujol, J. P. 1967. Le complex byssogène des mollusques bivalves. Histochimie comparée des sécrétions chez *Mytilus edulis* L. et *Pinna nobilis*. *Bull. Soc. Linn. Normandie* **8:** 308–332.

Rodrigues, V., M. Chandri, M. Knight, H. Meadows, A. E. Chambers, W. R. Taylor, C. Kelly, and A. J. G. Simpson. 1989. Predicted structure of a major *Schistosoma mansoni* eggshell protein. *Mol. Biochem. Parasitol.* **32:** 7–14.

Rzepecki, L. M., and J. H. Waite. 1991. Dopa proteins: versatile varnishes and adhesives from marine fauna. Pp. 119–148 in *Bioorganic Marine Chemistry*, Vol. 4, P. J. Scheuer, ed. Springer-Verlag, Berlin.

Rzepecki, L. M., S.-S. Chin, J. H. Waite, and M. F. Lavin. 1991. Molecular diversity of marine glues: polyphenolic proteins from five mussel species. *Mol. Mar. Biol. Biotechnol.* **1:** 78–88.

Rzepecki, L. M., K. M. Hansen, and J. H. Waite. 1992. Bioadhesives: DOPA and phenolic proteins as components of organic composite materials. In *Principles of Cell Adhesion*, P. D. Richardson, and M. Steiner, eds. CRC Press, Inc., Boca Raton, FL.

Segrest, J. P., and R. L. Jackson. 1972. Molecular weight determination of glycoproteins by polyacrylamide gel electrophoresis in sodium dodecyl sulfate. *Meth. Enzymol.* **28:** 54–63.

Tamarin, A., P. Lewis, and J. Askey. 1976. The structure and formation of the byssal attachment plaque-forming region in *Mytilus californianus*. *J. Morphol.* **149:** 199–221.

Tsugita, A., T. Uchida, H. W. Mewes, and T. Ataka. 1987. Rapid vapor phase acid hydrolysis of peptide and protein. *J. Biochem.* **102:** 1595–1597.

Vitellaro-Zuccarello, L. 1981. Ultrastructural and cytochemical study on the enzyme gland of the foot of a mollusc. *Tiss. Cell* **13:** 701–713.

Vitellaro-Zuccarello, L. 1980. The collagen gland of *Mytilus galloprovincialis*: an ultrastructural and cytochemical study on secretory granules. *J. Ultrastruct. Res.* **73:** 135–147.

Waite, J. H. 1987. Nature's underwater adhesive specialist. *Int. J. Adhesion Adhesives* **7:** 9–14.

Waite, J. H. 1990. The phylogeny and chemical diversity of quinone-tanned glues and varnishes. *Comp. Biochem. Physiol.* **97B:** 19–29.

Waite, J. H. 1991. Detection of peptidyl-DOPA by amino acid analysis and microsequencing techniques. *Anal. Biochem.* **192:** 429–433.

Waite, J. H. 1983. Evidence for a repeating 3,4-dihydroxyphenylalanine- and hydroxyproline-containing decapeptide in the adhesive protein of the mussel, *Mytilus edulis* L. *J. Biol. Chem.* **258:** 2911–2915.

Waite, J. H., D. C. Hansen, and K. T. Little. 1989. The glue protein of ribbed mussels (*Geukensia demissa*): a natural adhesive with some features of collagen. *J. Comp. Physiol.* **159B:** 517–525.

Waite, J. H., and A. C. Rice-Ficht. 1987. Presclerotized eggshell protein from the liver fluke *Fasciola hepatica*. *Biochemistry* **26:** 7819–7825.

Waite, J. H., and A. C. Rice-Ficht. 1989. A histidine-rich protein from the vitellaria of the liver fluke *Fasciola hepatica*. *Biochemistry* **28:** 6104–6110.

Waite, J. H., and M. L. Tanzer. 1981. Specific colorimetric detection of *o*-diphenols and 3,4-dihydroxyphenylalanine-containing peptides. *Anal. Biochem.* **111:** 131–136.